562
SCIENTIFIC AMERICAN OFFPRINTS

Eyewitness Testimony

by Robert Buckhout

SCIENTIFIC AMERICAN
DECEMBER 1974
VOL. 231, NO. 6 PP. 23-31

PUBLISHED BY **W. H. FREEMAN AND COMPANY** 660 MARKET STREET, SAN FRANCISCO, CALIFORNIA 94104

Copyright © 1974 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this offprint may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher.

Established 1845 SCIENTIFIC AMERICAN December 1974 Volume 231 Number 6

Eyewitness Testimony

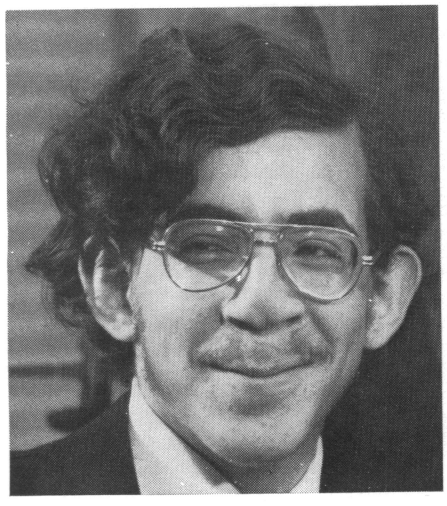
Although such testimony is frequently challenged, it is still widely assumed to be more reliable than other kinds of evidence. Numerous experiments show, however, that it is remarkably subject to error

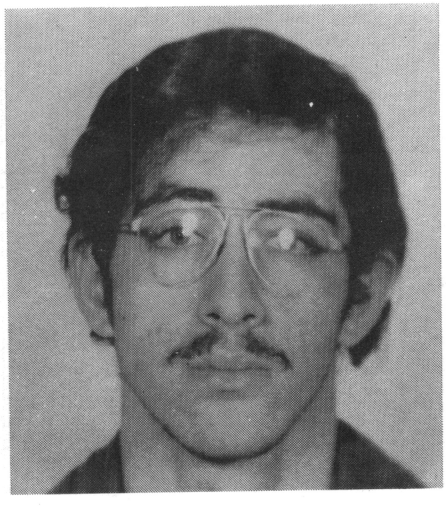
by Robert Buckhout

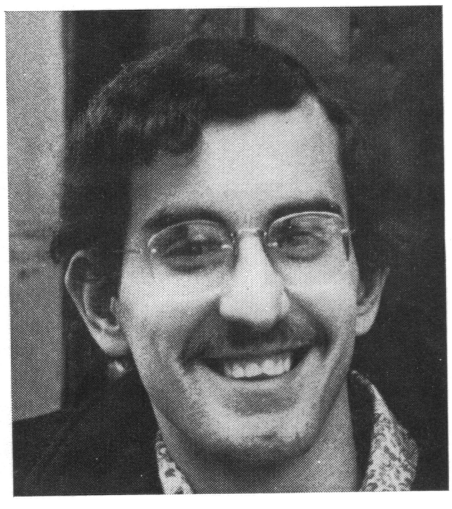
The woman in the witness box stares at the defendant, points an accusing finger and says, loudly and firmly, "That's the man! That's him! I could never forget his face!" It is impressive testimony. The only eyewitness to a murder has identified the murderer. Or has she?

Perhaps she has, but she may be wrong. Eyewitness testimony is unreliable. Research and courtroom experience provide ample evidence that an eyewitness to a crime is being asked to be something and do something that a normal human being was not created to be or do. Human perception is sloppy and uneven, albeit remarkably effective in serving our need to create structure out of experience. In an investigation or in court, however, a witness is often asked to play the role of a kind of tape recorder on whose tape the events of the crime have left an impression. The prosecution probes for stored facts and scenes and tries to establish that the witness's recording equipment was and still is in perfect running order. The defense cross-examines the witness to show that there are defects in the recorder and gaps in the tape. Both sides, and usually the witness too, succumb to the fallacy that everything was recorded and can be played back later through questioning.

Those of us who have done research in eyewitness identification reject that fallacy. It reflects a 19th-century view of man as perceiver, which asserted a parallel between the mechanisms of the physical world and those of the brain. Human perception is a more complex information-processing mechanism. So is memory. The person who sees an accident or witnesses a crime and is then asked to describe what he saw cannot call up an "instant replay." He must depend on his memory, with all its limitations. The limitations may be unimportant in ordinary daily activities. If someone is a little

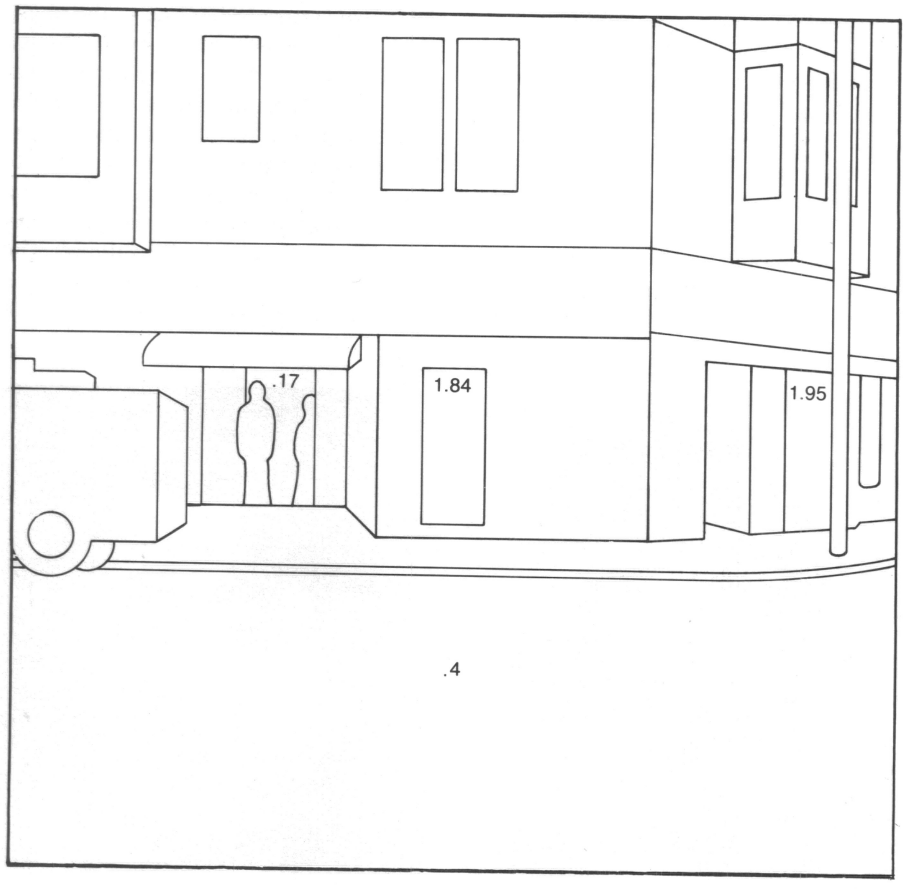






MISTAKEN IDENTIFICATIONS led to the arrests of two innocent men: Lawrence Berson (*left*) for several rapes and George Morales (*right*) for a robbery. Both men were picked out of police lineups by victims of the crimes. Berson was cleared when Richard Carbone (*center*) was arrested and implicated in the rapes. Carbone was convicted. Later he confessed to the robbery, clearing Morales.

REENACTMENT OF A MURDER was photographed at the same time of night as the murder and from the viewing position of an eyewitness who said he had been 120 feet away. The witness had identified a suspect charged with killing another man in darkened doorway.



MEASUREMENTS OF BRIGHTNESS (in millilamberts) at various points in the scene showed how little light was reflected from the doorway to the eyewitness. The light readings and the photograph (*top*) combined to cast doubt on the accuracy of the identification.

unreliable, if he trims the truth a bit in describing what he has seen, it ordinarily does not matter too much. When he is a witness, the inaccuracy escalates in importance.

Human perception and memory function effectively by being selective and constructive. As Ulric Neisser of Cornell University has pointed out, "Neither perception nor memory is a copying process." Perception and memory are decision-making processes affected by the totality of a person's abilities, background, attitudes, motives and beliefs, by the environment and by the way his recollection is eventually tested. The observer is an active rather than a passive perceiver and recorder; he reaches conclusions on what he has seen by evaluating fragments of information and reconstructing them. He is motivated by a desire to be accurate as he imposes meaning on the overabundance of information that impinges on his senses, but also by a desire to live up to the expectations of other people and to stay in their good graces. The eye, the ear and other sense organs are therefore social organs as well as physical ones.

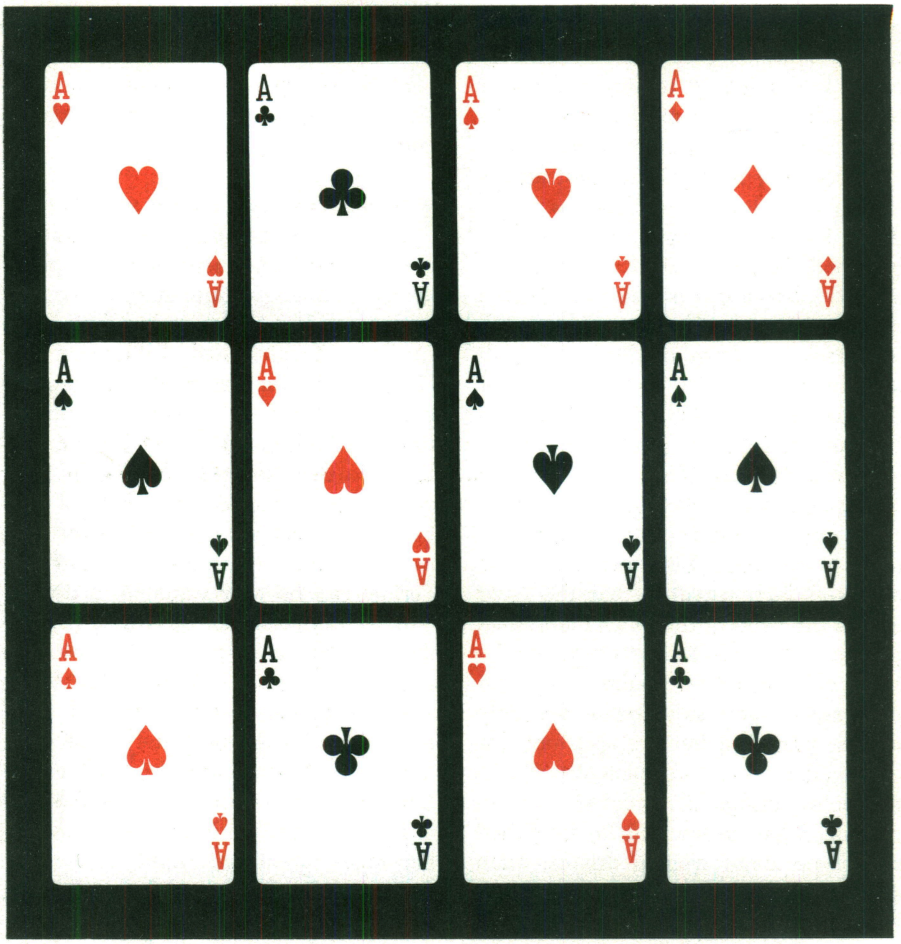
Psychologists studying the capabilities of the sense organs speak of an "ideal observer," one who would respond to lights or tones with unbiased eyes and ears, but we know that the ideal observer does not exist. We speak of an "ideal physical environment," free of distractions and distortions, but we know that such an environment can only be approached, and then only in the laboratory. My colleagues and I at the Brooklyn College of the City University of New York distinguish a number of factors that we believe inherently limit a person's ability to give a complete account of events he once saw or to identify with complete accuracy the people who were involved.

The first sources of unreliability are implicit in the original situation. One is the insignificance—at the time and to the witness—of the events that were observed. In placing someone at or near the scene of a crime, for example, witnesses are often being asked to recall seeing the accused at a time when they were not attaching importance to the event, which was observed in passing, as a part of the normal routine of an ordinary day. As long ago as 1895 J. McKeen Cattell wrote about an experiment in which he asked students to describe the people, places and events they had encountered walking to school over familiar paths. The reports were incomplete and unreliable; some individuals were

very sure of details that had no basis in fact. Insignificant events do not motivate a person to bring fully into play the selective process of attention.

The length of the period of observation obviously limits the number of features a person can attend to. When the tachistoscope, a projector with a variable-speed shutter that controls the length of an image's appearance on a screen, is used in controlled research to test recall, the shorter times produce less reliable identification and recall. Yet fleeting glimpses are common in eyewitness accounts, particularly in fast-moving, threatening situations. In the Sacco-Vanzetti case in the 1920's a witness gave a detailed description of one defendant on the basis of a fraction-of-a-second glance. The description must have been a fabrication.

Less than ideal observation conditions usually apply; crimes seldom occur in a well-controlled laboratory. Often distance, poor lighting, fast movement or the presence of a crowd interferes with the efficient working of the attention process. Well-established thresholds for the eye and the other senses have been established by research, and as those limits are approached eyewitness accounts become quite unreliable. In one case in my experience a police officer testified that he saw the defendant, a black man, shoot a victim as both stood in a doorway 120 feet away. Checking for the defense, we found the scene so poorly lit that we could hardly see a person's silhouette, let alone a face; instrument measurements revealed that the light falling on the eye amounted to less than a fifth of the light from a candle. The defense presented photographs and light readings to demonstrate that a positive identification was not very probable. The members of the jury went to the scene of the crime, had the one black juror stand in the doorway, found they could not identify his features and acquitted the defendant.



HOW MANY ACES OF SPADES DID YOU SEE? After a brief glance at this display of playing cards most people report seeing three. Actually there are five. Because people expect aces of spades to be black, not red, they tend to see only the black ones and to miss the atypical red ones. Thus do prior conditioning and experience influence perception.

The witness himself is a major source of unreliability. To begin with, he may have been observing under stress. When a person's life or well-being is threatened, there is a response that includes an increased heart rate, breathing rate and blood pressure and a dramatic increase in the flow of adrenalin and of available energy, making the person capable of running fast, fighting, lifting enormous weight—taking the steps necessary to ensure his safety or survival. The point is, however, that a person under extreme stress is also a less than normally reliable witness. In experimental situations an observer is less capable of remembering details, less accurate in reading dials and less accurate in detecting signals when under stress; he is quite naturally paying more attention to his own well-being and safety than to nonessential elements in the environment. Research I have done with Air Force flight-crew members confirms that even highly trained people become poorer observers under stress. The actual threat that brought on the stress response, having been highly significant at the time, can be remembered; but memory for other details such as clothing and colors is not as clear; time estimates are particularly exaggerated.

The observer's physical condition is often a factor. A person may be too old or too sick or too tired to perceive clearly, or he may simply lack the necessary faculty. In one case I learned that a witness who had testified about shades of red had admitted to the grand jury that he was color-blind. I testified at the trial that he was apparently dichromatic, or red-green color-blind, and that his testimony was probably fabricated in the basis of information other than visual evidence. The prosecution brought on his ophthalmologist, presumably as a rebuttal witness, but the ophthalmologist testified that the witness was actually monochromatic, which meant he could perceive no colors at all. Clearly the witness was "filling in" his testimony. That, after all, is how color-blind people function in daily life, by making inferences about colors they cannot distinguish.

Psychologists have done extensive research on how "set," or expectancy, is used by the observer to make judgments more efficiently. In a classic experiment done in the 1930's by Jerome S. Bruner and Leo Postman at Harvard University observers were shown a display of playing cards for a few seconds and asked to report the number of aces of spades in the display [*see illustration above*]. After a brief glance most observers reported seeing three aces of spades. Actually there were five; two of them were colored red instead of the more familiar black. People are so familiar with black aces of spades that they do not waste

time looking at the display carefully. The prior conditioning of the witness may cause him similarly to report facts or events that were not present but that he thinks should have been present.

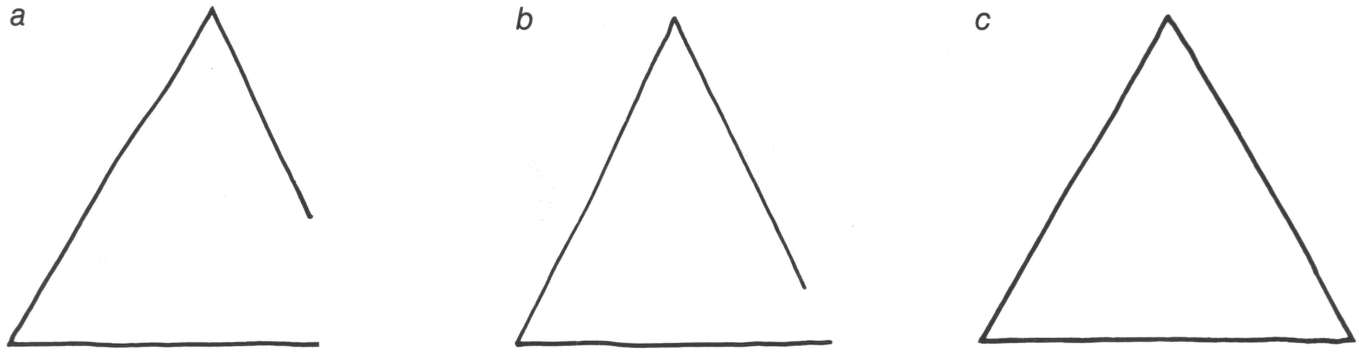
Expectancy is seen in its least attractive form in the case of biases or prejudices. A victim of a mugging may initially report being attacked by "niggers" and may, because of prejudice or limited experience (or both), be unable to tell one black man from another. ("They all look alike to me.") In a classic study of this phenomenon Gordon W. Allport of Harvard had his subjects take a brief look at a drawing of several people on a subway train, including a black man and a white man who is standing with a razor in his hand. Fifty percent of the observers later reported that the razor was in the hand of the black man. Most people file away some stereotypes on the basis of which they make perceptual judgments; such stereotypes not only lead to prejudice but are also tools for making decisions more efficiently. A witness to an automobile accident may report not what he saw but his ingrained stereotype about women drivers. Such short-cuts to thinking may be erroneously reported and expanded on by an eyewitness without his being aware that he is describing his stereotype rather than actual events. If the witness's biases are shared by the investigator taking a statement, the report may reflect their mutual biases rather than what was actually seen.

The tendency to see what we want or need to see has been demonstrated by numerous experiments in which people report seeing things that in fact are not present. R. Levine, Isador Chein and Gardner Murphy had volunteers go without food for 24 hours and report what they "saw" in a series of blurred slides presented on a screen. The longer they were deprived of food the more frequently they reported seeing "food" in the blurred pictures. An analysis of the motives of the eyewitness at the time of a crime can be very valuable in determining whether or not the witness is reporting what he wanted to see. In one study I conducted at Washington University a student dressed in a black bag that covered him completely visited a number of classes. Later the students in those classes were asked to describe the nature of the person in the bag. Most of their reports went far beyond the meager evidence: the bag-covered figure was said to be a black man, "a nut," a symbol of alienation and so on. Further tests showed that the descriptions were related to the needs and motives of the individual witness.

Journalists and psychologists have noted a tendency for people to maintain they were present when a significant historical event took place near where they live even though they were not there at all; such people want to sound interesting, to be a small part of history. A journalist once fabricated a charming human interest story about a naked woman stuck to a newly painted toilet seat in a small town and got it distributed by newspaper wire services. He visited the town and interviewed citizens who claimed to have witnessed and even to have played a part in the totally fictitious event. In criminal cases with publicity and a controversial defendant it is not uncommon for volunteer witnesses to come forward with spurious testimony.

Unreliability stemming from the original situation and from the observer's fallibility is redoubled by the circumstances attending the eventual attempt at information retrieval. First of all there is the obvious fact, supported by a considerable amount of research, that people forget verbal and pictorial information with the passage of time. They are simply too busy coping with daily life to keep paying attention to what they heard or saw; perfect recall of informa-

WHO HAD THE RAZOR? After a brief look at a drawing such as this one, half of the observers report having seen the razor, a stereotyped symbol of violence in blacks, in the black man's hand. Gordon W. Allport of Harvard University devised this experiment.



"FILLING IN" OF DETAILS was demonstrated by a simple drawing test. Observers were shown an incomplete but roughly triangular figure and immediately afterward were asked to draw what they had seen. The typical drawing was a good reproduction of the original (*a*). A month later observers asked to draw what they remembered produced more regular figures (*b*). Three months after the original viewing, again asked to draw what they remembered, they drew erroneously complete, symmetrical figures (*c*).

tion is basically unnecessary and is rarely if ever displayed. The testing of recognition in a police "lineup" or a set of identification photographs is consequently less reliable the longer the time from the event to the test. With time, for example, there is often a filling in of spurious details: an incomplete or fragmentary image is "cleaned up" by the observer when he is tested later. Allport used to have students draw a rough geometric shape right after such a shape was shown to them. Then they were tested on their ability to reproduce the drawing 30 days later and again three months later [*see illustration above*]. The observers tended first to make the figure more symmetrical than it really was and later to render it as a neat equilateral triangle. This finding was repeated with many objects, the tendency being for people to "improve" their recollection by making it seem more logical.

In analyses of eyewitness reports in criminal cases we have seen the reports get more accurate, more complete and less ambiguous as the witness moves from the initial police report through grand-jury questioning to testimony at the trial. The process of filling in is an efficient way to remember but it can lead to unreliable recognition testing: the witness may adjust his memory to fit the available suspects or pictures. The witness need not be lying; he may be unaware he is distorting or reconstructing his memory. In his very effort to be conscientious he may fabricate parts of his recall to make a chaotic memory seem more plausible to the people asking questions. The questions themselves may encourage such fabrication. Beth Loftus of the University of Washington has demonstrated how altering the semantic value of the words in questions about a filmed auto accident causes witnesses to distort their reports. When witnesses were asked a question using the word "smashed" as opposed to "bumped" they gave higher estimates of speed and were more likely to report having seen broken glass—although there was no broken glass.

Unfair test construction often encourages error. The lineup or the array of photographs for testing the eyewitness's ability to identify a suspect can be analyzed as fair or unfair on the basis of criteria most psychologists can agree on. A fair test is designed carefully so that all faces have an equal chance of being selected by someone who did not see the suspect; the faces are similar enough to one another and to the original description of the suspect to be confusing to a person who is merely guessing; the test is conducted without leading questions or suggestions. All too frequently lineups or photograph arrays are carelessly assembled or even rigged. If, for example, there are five pictures, the chance should be only one in five that any one picture will be chosen on the basis of guessing.

Frequently, however, one picture—the picture of the suspect—may stand out. In the case of the black activist Angela Davis one set of nine photographs used to check identification included three pictures of the defendant taken at an outdoor rally, two police "mug shots" of other women with their names displayed, a picture of a 55-year-old woman and so on. It was so easy for a witness to rule out five of the pictures as ridiculous choices that the test was reduced to four photographs, including three of Miss Davis. The probability was therefore 75 percent that a witness would pick out her picture whether he had seen her or not. Such a "test" is meaningless to a psychologist and is probably tainted as evidence in court.

Research on memory has also shown that if one item in the array of photographs is uniquely different—say in dress, race, height, sex or photographic quality—it is more likely to be picked out. Such an array is simply not confusing enough for it to be called a test. A teacher who makes up a multiple-choice test includes several answers that sound or look alike to make it difficult for a person who does not know the right answer to succeed. Police lineups and picture layouts are multiple-choice tests; if the rules for designing tests are ignored, the tests are unreliable.

No test, with photographs or a lineup, can be completely free of suggestion. When a witness is brought in by the police to attempt an identification, he can safely assume that there is some reason: that the authorities have a suspect in mind or even in custody. He is therefore under pressure to pick someone even if the officer showing the photographs is properly careful not to force the issue. The basic books on eyewitness identification all recommend that no suggestions, hints or pressure be transmitted to the witness, but my experience with criminal investigation reveals frequent abuse by zealous police officers. Such abuses include making remarks about which pictures to skip, saying, "Are you sure?" when the witness makes an error, giving hints, showing enthusiasm when the "right" picture is picked and so on. There is one version of the lineup in which five police officers in civilian clothes stand in the line, glancing obviously at the one real suspect. Suggestion can be subtler. In some experiments the test giver was merely instructed to smile and be very approving when a certain kind of photograph or statement was picked; such social approval led to an increase in the choosing of just those photographs even though there was no "correct" answer. A test that measures a need for social approval has shown that people who are high in that need

(particularly those who enthusiastically volunteer information) are particularly strongly influenced by suggestion and approval coming from the test giver.

Conformity is another troublesome influence. One might expect that two eyewitnesses—or 10 or 100—who agree are better than one. Similarity of judgment is a two-edged sword, however: people can agree in error as easily as in truth. A large body of research results demonstrates that an observer can be persuaded to conform to the majority opinion even when the majority is completely wrong. In one celebrated experiment, first performed in the 1950's by Solomon E. Asch at Swarthmore College, seven observers are shown two lines and asked to say which is the shorter. Six of the people are in the pay of the experimenter; they all say that the objectively longer line is the shorter one. After hearing six people say this, the naïve subject is on the spot. Astonishingly the majority of the naïve subjects say that the long line is short—in the face of reality and in spite of the fact that alone they would have no trouble giving the correct answer [see "Opinions and Social Pressure," by Solomon E. Asch; SCIENTIFIC AMERICAN, November, 1955].

To test the effect of conformity a group of my students at Brooklyn College, led by Andrea Alper, staged a "crime" in a classroom, asked for individual descriptions and then put the witnesses into groups so as to produce composite descriptions of the suspect. The group descriptions were more complete than the individual reports but gave rise to significantly more errors of commission: an assortment of incorrect and stereotyped details. For example, the groups (but not the individuals) reported incorrectly that the suspect was wearing the standard student attire, blue jeans.

The effects of suggestion increase when figures in obvious authority do the testing. In laboratory research we find more suggestibility and changing of attitudes when the tester is older or of apparently higher status, better dressed or wearing a uniform or a white coat—or is a pretty woman. In court I have noticed that witnesses who work together under a supervisor are hard put to disagree with their boss in testifying or in picking a photograph. The process of filling in details can be exaggerated when the boss and his employee compare their information and the employee feels obligated to back up his boss to remain in his good graces. Legal history is not lacking in anecdotes about convict witnesses who were rewarded by the authorities for their cooperation in making an identification.

In criminal investigations, as in scientific investigations, a theory can be a powerful tool for clarifying confusion, but it can also lead to distortion and unreliability if people attempt, perhaps unconsciously, to make fact fit theory and close their minds to the real meanings of facts. The eyewitness who feels pressed to say something may shape his memory to fit a theory, particularly a highly publicized and seemingly reasonable one. Robert Rosenthal of Harvard studied this effect. He devised a test in

TWO LAYOUTS OF PHOTOGRAPHS like these were presented to witnesses to a staged assault. The actual attacker had been the young man labeled No. 5. In the unbiased spread (*left*) the portraits are aligned and show similar full-face views; in the biased spread (*right*) the culprit's head is tilted and he is grinning, and the portrait itself is placed at an angle. Witnesses were shown one

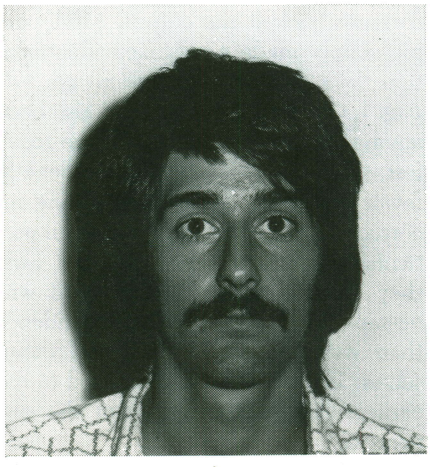
which people were supposed to pick out a "successful" face from a set of photographs. There was actually no correct answer, but the experimenter dropped hints to his assistants as to what he thought the results should be. When they subsequently administered the test the assistants unconsciously signaled the subjects as to which photograph to pick, thus producing results that supported their boss's theory. Any test is a social interaction as well as a test.

There is a nagging gap between data on basic perceptual processes in controlled research settings and important questions about perception in the less well-controlled real world. Inspired by the new approach to perception research exemplified in the work of Neisser and of Ralph Norman Haber of the University of Rochester, my colleagues and I have felt that this gap can only be bridged by conducting empirical research on eyewitness identification in a somewhat real world. In one such experiment we staged an assault on the campus of the California State University at Hayward: a student "attacked" a professor in front of 141 witnesses; another outsider of the same age was on the scene as a bystander. We recorded the entire incident on videotape so that we could compare the true event with the eyewitness reports. After the attack we took sworn statements from each witness, asking them to describe the suspect, his clothes and whatever they could remember about the incident. We also asked each witness to rate his own confidence in the accuracy of his description.

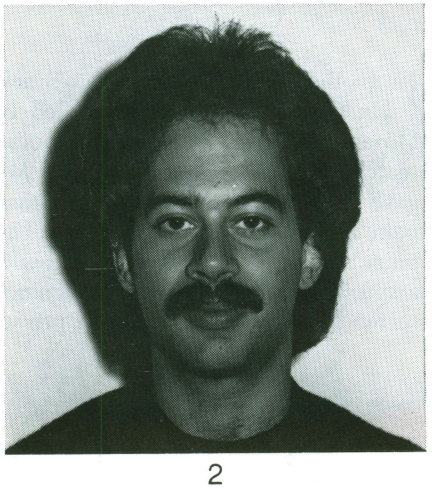
As we expected, the descriptions were quite inaccurate, as is usually the case in such situations. The passage of time was overestimated by a factor of almost two and a half to one. The average weight estimate for the attacker was 14 percent too high, and his age was underestimated by more than two years. The total accuracy score, with points given for those judgments and for others on appearance and dress, was only 25 percent of the maximum possible score. (Only the height estimate was close. This may be because the suspect was of average height; people often cite known facts about the "average" man when they are uncertain.)

We then waited seven weeks and presented a set of six photographs to each witness individually under four different experimental conditions. There were two kinds of instructions: low-bias, in which witnesses were asked only if they recognized anybody in the photographs and high-bias, in which witnesses were reminded of the attack incident, told that we had an idea who the suspect was and asked to find the attacker in one of two arrangements of photographs, all well-lit frontal views of young men including the attacker and the bystander. In the unbiased picture spread all six portraits were neatly set out with about the same expression on all the faces and with similar clothing. In the biased spread the attacker was shown with a distinctive expression and his portrait was positioned at an angle [*see illustration on these two pages*].

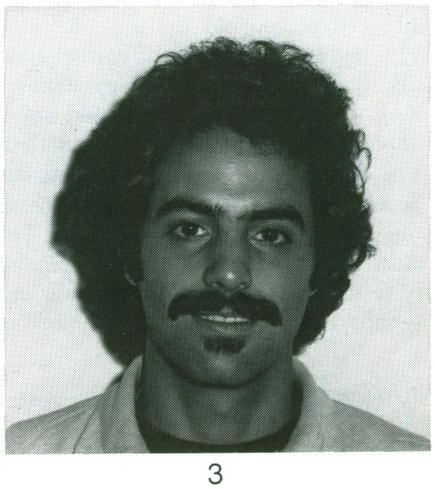
Only 40 percent of the witnesses iden-



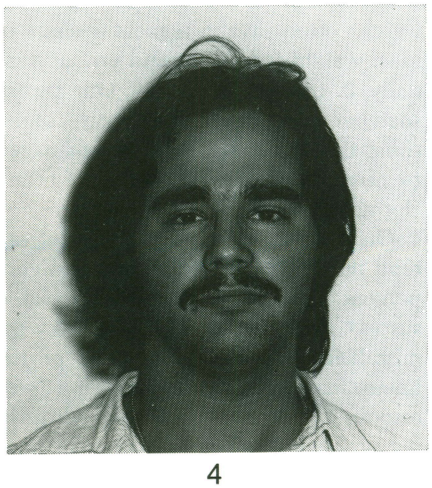
1



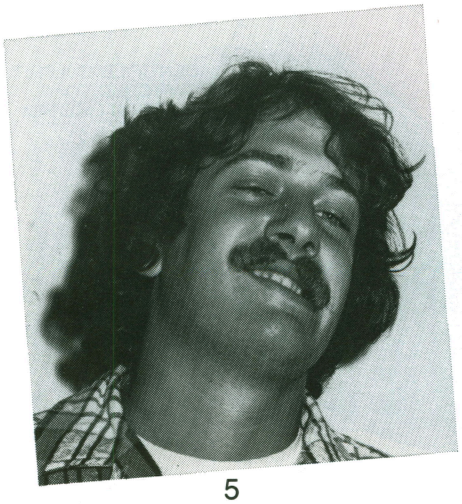
2



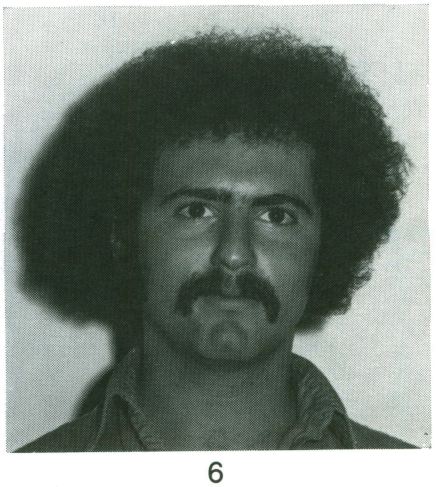
3



4

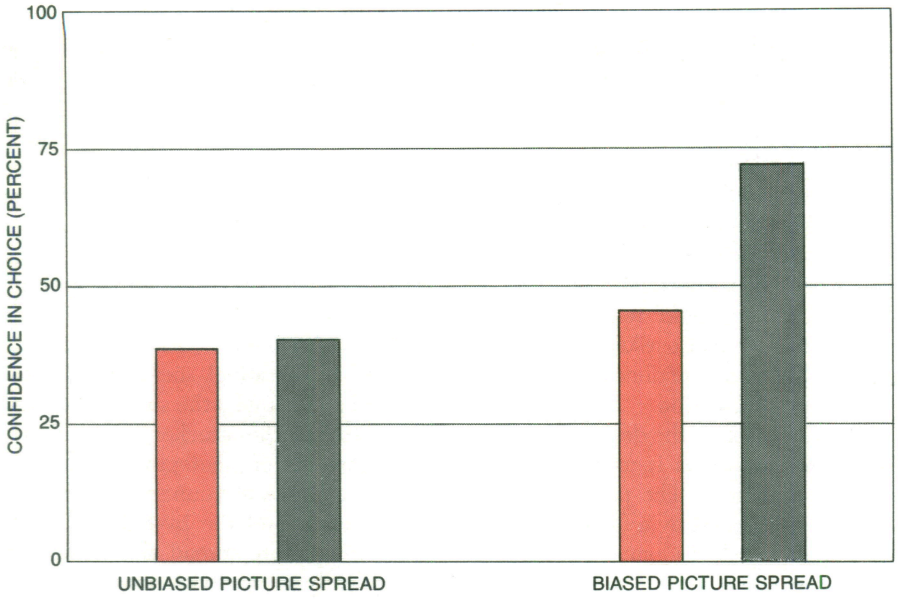


5



6

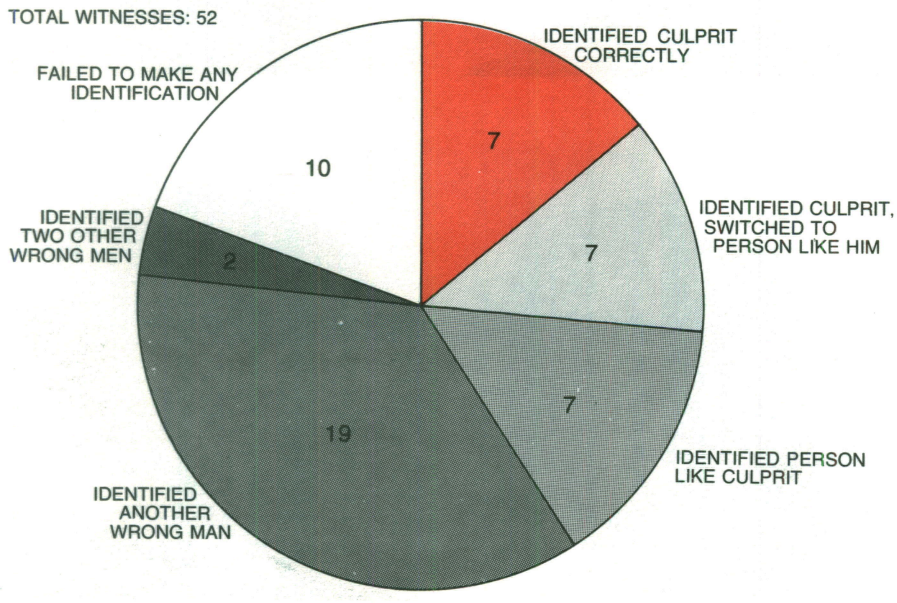
of the spreads after having been given one of two kinds of instructions: either low-bias (simply, "Do you recognize any of these men?") or high-bias (such as, "One of these men is a suspect in that assault we saw; it is important that you identify him for us"). Whereas 40 percent of all the witnesses picked No. 5, 61 percent of those who saw the biased spread and got biased instructions did so.



BIASED CONDITIONS gave observers more confidence in their ability to recognize faces in the picture layouts displayed on the two preceding pages. The bars show the degree of confidence expressed by those who picked from the unbiased spread (*left*) and the biased one (*right*) and after having been given low-bias (*color*) and high-bias (*gray*) instructions.

tified the suspect correctly; 25 percent of them identified the innocent bystander instead; even the professor who was attacked picked out the innocent man. The highest proportion of correct identifications, 61 percent, was achieved with a combination of a biased set of photographs and biased instructions. The degree of confidence in picking suspect No. 5, the attacker, was also significantly higher in that condition [*see illustration above*]. We have subsequently tested the same picture spreads with groups that never saw the original incident. We describe the assault and ask people to pick the most likely perpetrator. Under the biased conditions they too pick No. 5.

In another study undertaken at Brooklyn College a student team, led by Miriam Slomovits, staged a live purse-snatching incident in a classroom. We gave the witnesses the usual questionnaire and got the usual bad scores. This time, however, we were concerned with a specific dilemma: Why is recognition so much better than recall? In private



REAL CULPRIT in a staged purse-snatching incident was identified by only seven of 52 witnesses who viewed two videotaped lineups. Seven others picked the culprit first but then switched to a man who looked like him in the second lineup; seven picked only the man who looked like him. Most witnesses picked other people or were unable to choose.

most lawyers and judges agree that the recall of a crime by a witness is very bad, but they still believe people can successfully identify a suspect. What we had to do was to break away from our demonstrations of how bad witnesses are at recalling details and search for what makes a witness good at recognizing a face. To do so we took the witnesses who had predictably given poor recall data and gave them a difficult recognition test. Our witnesses got not only a lineup with the actual purse-snatcher in the group but also a second lineup that included only a person who looked like the purse-snatcher. The question was: Would the witnesses pick only the real culprit and avoid making a mistaken identification of the person who looked like him?

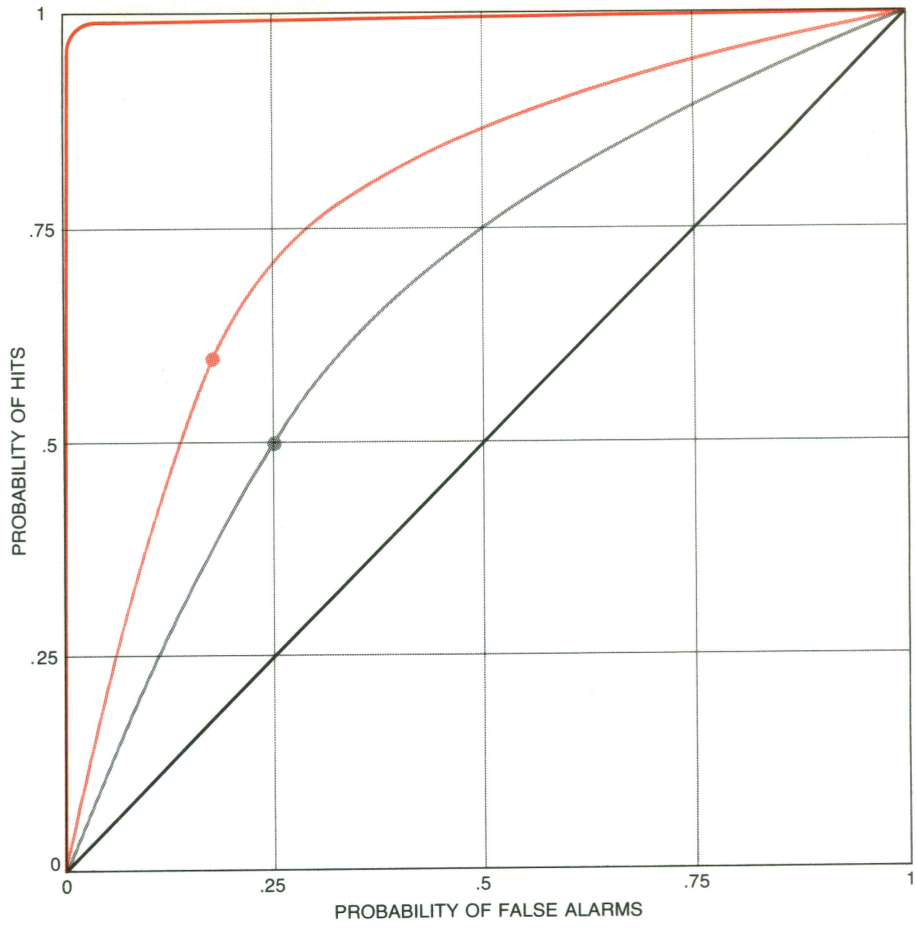
We videotaped two lineups of five persons each and showed them in counterbalanced order to 52 witnesses of the purse-snatching. Very few witnesses were completely successful in making a positive identification without ambiguity. An equal number of witnesses impeached themselves by picking the man who resembled the culprit after having correctly picked the culprit. Most people simply made a mistaken identification [*see bottom illustration on this page*]. Our best witnesses had also been among the best performers in the recall test, that is, they had made significantly fewer errors of commission (adding incorrect details). They had not given particularly complete reports, but at least they had not filled in. The good witnesses also expressed less confidence than witnesses who impeached themselves. Finally, when we referred to the earlier written descriptions of the suspect we found our successful witnesses had given significantly higher, and hence more accurate, estimates of weight. People guessing someone's weight often invoke a mental chart of ideal weight for height and err substantially if the person is fat. Our purse-snatcher was unusually heavy, something the successful witnesses managed to observe in spite of his loose-fitting clothing. The others were guessing.

Once again we noted that witnesses tend not to say, "I don't know." Eighty percent of our witnesses tried to pick the suspect even though most of them were mistaken. The social influence of the lineup itself seems to encourage a "yes" response. This effect presented a disturbing problem that actually drove us back from these rather realistically enacted crimes to the more controlled, emotionally neutral environment of the laboratory. We hoped to design a test for eye-

witnesses that could distinguish a good witness from a poor one, under circumstances in which we knew what the true facts were.

Pure measures of accuracy would not be adequate, since there are many different kinds of error, some of which come from the witness's desire to please the questioner with an abundance of details. Eventually we settled on adapting signal-detection theory, espoused by John A. Swets of the Massachusetts Institute of Technology, to the eyewitness situation. Signal-detection theory evolved in psychophysics as a means of coping with the fact that an observer's attitude "interferes" with his detection, processing and reporting of sensory stimuli. Limited to saying "yes" or "no" (I hear or see or smell it, or whatever), the observer applies criteria that vary with personality, experience, anticipated cost or reward, motivation to please the tester or to frustrate him and other factors. What the experimenter does, therefore, is usually to present noise about half of the time and signals plus noise about half of the time and to count correct "yes" answers (hits) and incorrect "yes" answers (false alarms), combining the scores statistically into a single measure of observer sensitivity. This quantifies an estimate of the observer's criteria for judging his immediate experience. A very cautious person might have very few false alarms and a high proportion of hits, indicating that he says "yes" sparingly; a less than cautious person might say "yes" most of the time, scoring a large number of hits but only at the price of a large number of false alarms.

In our research at Brooklyn College Lynne Williams and I now show a film of a supposed crime and then present to the observers 20 true statements about the incident and the same number of false statements. The witness indicates "yes" or "no" as to the truth of each statement. We end up with a record of hits and false alarms which, after some complicated statistical processing, yields a curve called a receiver-operating-characteristic (ROC) curve [*see illustration on this page*]. A person whose hits and false alarms were equal, indicating that the answers had no relation to the true facts, would generate a straight diagonal ROC curve. A perfect witness would have all hits and no false alarms. Real people fall somewhere in between. We have found so far that witnesses with the better (which is to say higher) ROC curves go on to do better than other people at recognizing the suspect in a lineup. We are using the ROC function to test various hypotheses about how environmental conditions, stress, mental set, bias in interrogation, age, sex, and social, ethnic and economic group affect the accuracy and reliability of eyewitnesses.



PERFORMANCE AS A WITNESS is measured by plotting "hits" against "false alarms." Hits are "yes" answers regarding the truth of a true descriptive detail of a scene the witness has viewed; false alarms are "yes" answers regarding the truth of a false detail. "Witness sensitivity curves" are generated by the results for various witnesses or for answers given by the same witness with varying degrees of confidence. A perfect witness would be one who always scored hits and never scored false alarms (*solid color curve*); a "blind" witness would score as many hits as false alarms (*black*). In the author's experiment successful witnesses, who identified a suspect correctly in a lineup, had produced curves that were higher on the chart; they averaged 12 hits to 3.6 false alarms (*light color*). Unsuccessful witnesses had produced lower curves, averaging 10 hits to five false alarms (*gray*). Scores are plotted as fractions of the maximum possible scores: 20 hits or 20 false alarms.

Psychological research on human perception has advanced from the 19th-century recording-machine analogy to a more complex understanding of selective decision-making processes that are more human and hence more useful. My colleagues and I feel that psychologists can make a needed contribution to the judicial system by directing contemporary research methods to real-world problems and by speaking out in court (as George A. Miller of Rockefeller University puts it, by "giving psychology away").

It is discouraging to note that the essential findings on the unreliability of eyewitness testimony were made by Hugo Münsterberg nearly 80 years ago, and yet the practice of basing a case on eyewitness testimony and trying to persuade a jury that such testimony is superior to circumstantial evidence continues to this day. The fact is that both types of evidence involve areas of doubt. Circumstantial evidence is tied together with a theory, which is subject to questioning. Eyewitness testimony is also based on a theory, constructed by a human being (often with help from others), about what reality was like in the past; since that theory can be adjusted or changed in accordance with personality, with the situation or with social pressure, it is unwise to accept such testimony without question. It is up to a jury to determine if the doubts about an eyewitness's testimony are reasonable enough for the testimony to be rejected as untrue. Jurors should be reminded that there can be doubt about eyewitness testimony, just as there is about any other kind of evidence.

The Author

ROBERT BUCKHOUT is associate professor of psychology and director of the Center for Responsive Psychology at Brooklyn College of the City University of New York. "I was born in Brooklyn," he writes, "and after leaving with a vow never to return, I am back—minutes away from where I was born. My career as a psychologist began as a major subject chosen after minoring in mistakes. After taking a master's degree at North Carolina State University I spent four years in the Air Force. I went to Ohio State University, obtaining my Ph.D. In 1963 I began teaching at Washington University and getting involved in the social problems of the time." Buckhout adds that he owes "more than I can express in words" to a group of undergraduates who helped him with the research on which his article is based. They include Andrea Alper, Susan Chern, Robinsue Frohboese, Richard Harwood, Daryl Figueroa, Ethan Hoff, Carolyn Hogan, Noreen Norton, Vincent Reilly, Betti Sachs, Glen Silverberg, Miriam Slomovits and Lynne Williams. He writes that he is "also indebted to Leo Branton and Howard Moore, attorneys, who taught me to present psychology in English to a jury."

Bibliography

ON THE WITNESS STAND: ESSAYS ON PSYCHOLOGY AND CRIME. Hugo Münsterberg. Doubleday, Page and Company, 1915.

EYEWITNESS IDENTIFICATION IN CRIMINAL CASES. Patrick M. Wall. Charles C. Thomas, Publisher, 1965.

EXPERIMENTOR EFFECTS IN BEHAVIORAL RESEARCH. Robert Rosenthal. Appleton-Century-Crofts, 1966.

THROUGH A BAG, DARKLY, Robert Buckhout in *American Psychologist*, Vol. 23, No. 11, pages 832–833; November, 1968.